AGILE MINING
- A NOVEL DATA MINING PROCESS FOR INDUSTRY PRACTICE BASED ON AGILE METHODS AND VISUALIZATION

Xiao Zhu

Advanced Analytics Institute

School of Software

Faculty of Engineering and Information Technology

University of Technology Sydney

This dissertation is submitted for the degree of Master by Research

November 2017

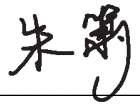
Publication List during Master Degree Study:

X. Zhu, G. Xu. Applying Visual Analytics on Traditional Data Mining Process: Quick Prototype, Simple Expertise Transformation, and Better Interpretation. *International Conference on Enterprise Systems: Advances in Enterprise Systems* (ES), 2016.

S. Liu, G. Xu, X. Zhu. Towards Simplified Insurance Application via Sparse Questionnaire Optimization. *International Conference on Behavioral, Economic, and Socio-Cultural Computing* (BESC), 2017.

DECLARATION

This dissertation is the result of my work and includes nothing, which is the outcome of work done in collaboration except where specifically indicated in the text. It has not been previously submitted, in part or whole, to any university or institution for any degree, diploma, or other qualification.

Signed: 

Date: 15/11/2017

Xiao Zhu

ABSTRACT

Current standard data mining processes like CRoss-Industry Standard Process for Data mining (CRISP-DM) are vulnerable to frequent change of customer requirement. Meanwhile, Stakeholders might not acquire sufficient understanding to generate business value from analytic results due to a lack of intelligible explanatory stage. These two cases repeatedly happen on those companies which are inexperienced in data mining practice. Towards this issue, Agile Mining, a refined CRISP-DM based data mining (DM) process, is proposed to address these two friction points between current data mining processes and inexperienced industry practitioners. By merging agile methods into CRISP-DM, Agile Mining process achieves a requirement changing friendly data mining environment for inexperienced companies. Moreover, this Agile Mining transforms traditional analytic-oriented evaluation to business-oriented visualization-based evaluation. In the case study, two industrial data mining projects are used to illustrate the application of this new data mining process and its advantages.

Keyword: Data Mining, Data Mining Standard Process, CRISP-DM, Visualization.

CONTENTS

List of Abbreviations and Acronyms

CRISP-DM: Cross Industry Standard Process for Data Mining

AM: Agile Mining

RR: Requirement Reconfirming

PR: Performance Refining

RF: Random Forest

LR: Logistic Regression

SVM: Support Vector Machine

NB: Naïve Bayes